Sentiment Analysis on the Fake News Spread in Twitter during 2020 US Presidential Elections

Sampreeth Avvari* **Barath Rama Shankar*** **Dhruv Sridhar***
spa9659@nyu.edu, br2543@nyu.edu, ds7395@nyu.edu
New York University

Abstract

Political sentiment analysis reveals trends in digital discourse where users often express support for candidates by criticizing opponents. This reflects a dynamic where backing is shown more through opposition rather than positive endorsement. Our study analyzes millions of tweets to understand the polarized emotions and attitudes during elections. We use a **fine-tuned Few-shot Llama 3** model to detect fake news and a **pretrained RoBERTa** model to categorize tweets as pro-candidate or anti-candidate based on sentiment. Additionally, **LDA** helps identify key topics within these tweets and assess their impact on different states by comparison with public surveys. This research evaluates advanced deep learning techniques for fake news detection and examines Twitter's content moderation for potential partisan bias, focusing on tweets related to **#JoeBiden** and **#DonaldTrump**. The model and associated code accessible via the **GitHub repository**.

Introduction

The influence of social media on political discourse is significant, especially during events like the U.S. presidential elections. Social media offers a direct insight into public sentiment, where users express support or disdain for candidates through tweets. This study analyzes tweets from the 2020 U.S. presidential election using **#JoeBiden** and **#DonaldTrump** to assess claims of partisan bias in Twitter's moderation and the spread of misinformation. By categorizing tweets as True, False, and Personal Opinion, and analyzing sentiment, we explore the intensity and nature of political emotions during the election season.

The key objectives of our research are outlined as follows:

- **Model Optimization:** Fine-tune the Few-shot Llama 3 model for detecting fake news in U.S. elections.
- **Sentiment Classification:** Use RoBERTa to classify tweets by sentiment into pro or anti-candidate categories.
- **Topics Modeling:** Examine election season topics and their variations with sentiment changes.
- **Misinformation Mapping:** Measure fake news spread across states and partisans in 2020, assessing geographical and temporal trends.
- **Geographic Distribution:** Study regional sentiment distribution to gauge localized political engagement.
- **Behavioral Insights:** Explore user patterns in political support, focusing on antagonistic versus supportive expressions in misinformation-laden tweets.

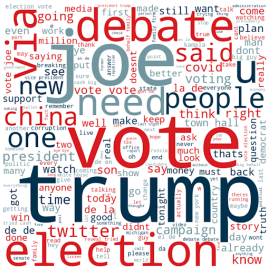

Figure 1: Word Cloud of 2020 Election Tweets

Literature Survey

The question of political bias in social media, has been a contentious topic with varying perspectives and claims. Research and public debate on this issue reflect a complex interplay of user experiences, algorithmic influences, and changes in platform management. Over the years, sentiment analysis has evolved from simple word dictionaries to advanced machine learning models.

Summary of Key Studies and Their Takeaways

- **Knight Foundation Study [1]:** Quantitative analysis of over 10 million tweets to identify misinformation spread by bots during the 2016 U.S. presidential election. ***Key Takeaway:*** Bots significantly influenced the dissemination of misinformation, with few accounts contributing to the majority of misleading content.
- **Identifying Tweets with Fake News IEEE Paper [2]:** Development of a classification system to distinguish real and fake tweets using feature extraction techniques. Key Takeaway: Effective identification and classification of

References:

[1] Knight Foundation, "Misinformation," 2021. https://knightfoundation.org/features/misinfo/

[2] Y. M. Said, Z. Ibrahim, and N. A. M. Ali, "Autonomic computing: A fuzzy control approach," in Proc. IEEE Int. Conf. on Computer and Communication Engineer-ing, Kuala Lumpur, 2018, pp. 1-6, doi: 10.1109/IC-CCE.2018.8424744.

fake news can be achieved through advanced feature extraction and data mining techniques.

- **Sentiment Analysis Using MLP Classifier [3]:** Analyzed tweet sentiments using a Multi-Layer Perceptron (MLP) classifier, initially trained on a limited dataset. ***Key Takeaway:*** Highlighted a balanced sentiment distribution but underscored potential limitations due to initial training scope.
- **Sentiment Analysis with VADER [4]:** Applied the VADER tool to 7.6 million tweets, including posts from accounts that were subsequently deleted or suspended. ***Key Takeaway:*** Demonstrated VADER's utility in detecting subtle shifts in sentiment and language nuances on social media.

The study, although thorough, relies on key assumptions:

Philosophical Assumptions

1. **Content Moderation Bias:** Suggests platform bias if misinformation removal is uneven between parties, indicating possible bias at the platform level.
2. **Impartiality Challenges:** The difficulty in proving platform bias based on tweet classifications with the resources and models available leaves certain questions about platform behavior and content moderation open.

Technical Assumptions

1. **Tweet Classification Validity:** Despite the challenges posed by the large dataset, it is assumed that hashtags accurately reflect discussions about the candidates.
2. **Topic Modeling Effectiveness:** Expects topic modeling to reveal representative political discourse themes.
3. **Misinformation Detection:** Utilizes a fine-tuned Few-shot Llama 3 model on the "Liar" dataset to identify misinformation spread and ensure content accuracy.
4. **Model Capability and Neutrality:** Assumes the Few-shot Llama 3 model is capable and unbiased in categorizing tweets accurately.

Methodology

Suggesting the integration of Large Language Models (LLMs) to more efficiently and insightfully analyze tweet authenticity, providing a scalable and cost-effective solution for social media platforms.

I. Datasets

1. LIAR Dataset [5]: The LIAR dataset, introduced by Wang et al., is a significant resource for fake news detection research. It comprises 12,800 manually labeled short statements derived from `POLITIFACT.COM`, spanning a decade and analyzed for truthfulness by professional editors. The dataset includes details such as the statement, speaker, context, and class label(refer Fig 2). It is segmented into three parts: 10,000 statements for training and approximately 2,000 statements each for validation and testing. The dataset has the following Class Labels: True, Mostly True, Half True, Barely True, Flase, Pants Fire. Initially, we transformed the dataset into a format compatible with Llama-3.

PreProcessing: Specifically, we processed the statement, output, and added an instructional column to serve as the system prompt for the LLM. To simplify the classification task, we consolidated the original five labels into two categories: true and false.

2. US 2020 Election Tweets Dataset[6]: This dataset was compiled to analyze public sentiment on Twitter for the 2020 U.S. Presidential Election. It focuses on the relationship between Twitter discourse and election outcomes, emphasizing election manipulation, candidate support prediction, and the geographic distribution of discussions. Covering October 15 to November 8, 2020, it includes approximately 1.7 million tweets tagged with #JoeBiden and #DonaldTrump.

PreProcessing: Given the prevalence of emojis and other non-text characters in tweets, preprocessing involved converting emojis to text by removing emoji characters. Next, removing links, hashtag symbols, and other non-ASCII characters while preserving the base textual information. We dropped a irrelevant features and used only `created_at`, tweet, number of likes, retweets, state, and country.

II. Models

1. Fake News Model:
We employed two distinct models for inference. The first, LLAMA-3 8B [7], from Meta's suite of LLMs, uses a tokenizer with a vocabulary of 128,000 tokens, enhancing its text understanding. It was fine-tuned on the LIAR dataset to detect misinformation. This fine-tuning process adapted the model to the nuances of deceptive language, enhancing its precision in real-world scenarios. LLAMA-3 8B was fine-tuned using Unsloth AI's QLora adapters, optimized with 4-bit quantization for NVIDIA's GPUs, suitable for high-performance machine learning tasks.

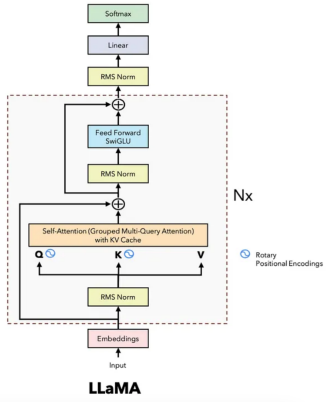

Figure 2: LLaMA Architecture

Training Details of LLaMA 3:

- **Training Duration:** 30 epochs.
- **Framework Used:** Unsloth's framework, optimizing GPU usage.
- **Efficiency Boost:** Achieved up to 30x faster training.

[3] H. Liu, H. Yue, and E. Xia, "Tweet Sentiment Analysis of the 2020 U.S. Presidential Election," in Proc. of The Web Conf. 2021, Ljubljana, Slovenia, Apr. 2021, doi: 10.1145/3442442.3452322

[4] Ali, R.H., et al., "A large-scale sentiment analysis of tweets pertaining to the 2020 US presidential election," J Big Data, vol. 9, no. 79, 2022. doi: 10.1186/s40537-22-00633-z

[5] Liar dataset, Hugging Face. https://huggingface.co/datasets/liar

[6] M. Hui, "US Election 2020 Tweets," Kaggle. https://www.kaggle.com/datasets/manchunhui/us-election-2020-tweets

[7] Meta AI, "Introducing Meta LLAMA 3," Meta AI Blog. https://ai.meta.com/blog/meta-llama-3/

- **Hyperparameter Tuning:** Initial learning rate set at 2e-4, optimized to 8e-6.
- **Optimizer:** Used the ”adamw_8bit” optimizer.
- **Logging:** Set logging_steps to 1 for frequent updates.
- **Inference:** We utilized a 12-shot learning approach with our fine-tuned Llama-3 model for Fake News Detection. This approach significantly improved the model's ability to recognize contextually relevant factors in tweets potentially containing misinformation. By training with a strategically selected set of examples, the model quickly adapted to subtle cues and patterns typical of false information. Despite GPU resource limitations, we analyzed 155,000 tweets, efficiently classifying content veracity and detecting misinformation without large datasets.

2. Sentiment Analysis Model

i. RoBERTa:

The RoBERTa-base sentiment analysis model [8], developed under the TimeLMs project, is a refined natural language processing tool, trained on approximately 124 million tweets from January 2018 to December 2021. It was fine-tuned using the TweetEval benchmark, a standard for evaluating Twitter models, enabled it to adeptly handle tweets' informal and distinctive language. RoBERTa utilizes a architecture that processes words via self-attention, allowing simultaneous consideration of all words in a sentence, providing a comprehensive context, leading to a deeper understanding of the text. The 'base' version used in this study includes 12 transformer layers, with each layer having 768 hidden units and 12 self-attention heads, optimizing it for the complexities of social media language.

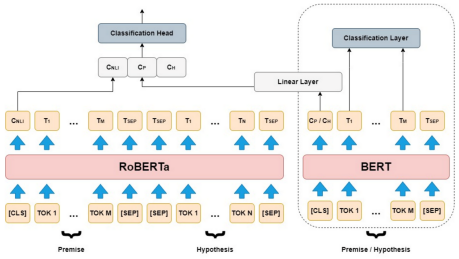

Figure 3: RoBERTa Architecture

ii. LDA:

Latent Dirichlet Allocation (LDA) [9] is a widely used generative statistical model for topic modeling and document classification. It assumes each document is a mixture of various latent topics, where each topic is a distribution of words. LDA starts by randomly assigning topic representations to documents, then iteratively refines these topics based on the words in the documents through methods like Gibbs sampling or variational Bayes, until the topics stabilize. This process effectively characterizes documents by a distribution of topics and topics by a distribution of words.

Inference: We utilized a zero-shot learning approach with the Twitter-RoBERTa-base model specifically tailored for sentiment analysis. This helped to analyze sentiment into positive, negative or neutral. To delve deeper into the thematic structures of the tweets, we employed the LDA model for topic modeling. This method facilitated the discovery of prevalent topics across the dataset, providing insights into the primary concerns and themes discussed during the elections.

3. Overall: Integrating fake news detection, sentiment analysis, and topic modeling enabled us to map sentiment shifts and misinformation spread related to specific topics during the election period, offering insights into public sentiment dynamics, the spread of misinformation and its impact during the election period.

Results

The analysis shows a significant presence of both fake and negative tweets about Trump during the peak election season, with his name appearing in approximately in 52% of the tweets compared to just 4% mentioning ”Joe” or ”Biden”. This disparity underscores the extent of discussion surrounding Trump, with the majority (60%) of the mentions being negative. ”Trump” trumped over ”Joe”!

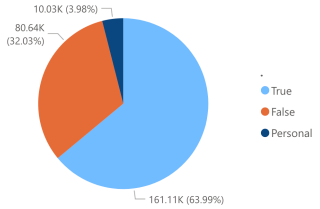

Figure 4: Overall Fake News Spread - The spread of fake news is around 32% in all the tweets, showcasing a large amount of false statements spread across social media.

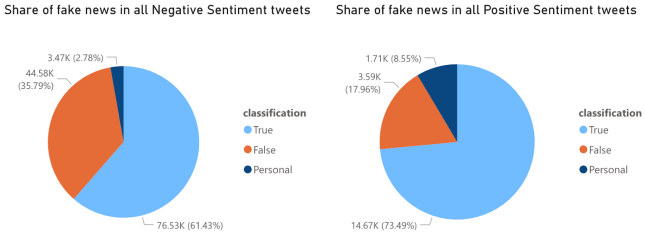

Figure 5: Overall Fake News Spread - Fake News spread in Positive News - 18% (Right) is significantly less (almost Half) than that in Negative news - 36% (Left)

[8] CardiffNLP, "Twitter-roBERTa-base-sentiment-latest," Hugging Face. https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment-latest

[9]T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, and J. Brew, "Huggingface's Transformers: State-of-the-art Natural Language Processing," arXiv, 2017 https://arxiv.org/pdf/1711.04305

Table 1: Topic Analysis based on Classification and Sentiment

S.No.	Classification	Sentiment	Hashtag	Topics	Topic Analysis
1	True	Positive	#JoeBiden	Love and Support	Emotionally charged words like ”kind,” ”love” and ”support” indicates that voters were encouraged to vote for Biden.
2	True	Positive	#DonaldTrump	Greatest President	Appearance of positive terms like ”good,” ”great,” ”thanks,” ”trump” reflects a tone of positivity and gratitude.
3	True	Negative	#JoeBiden	Covid and Trump	The words ”Narcissist,” ”Racist,” and ”Trump” in #JoeBiden tweets suggest that Trump was a central focus of the discussion.
4	True	Negative	#DonaldTrump	Pandemic and Taxes	Words like ”worst,” ”tax,” ”liar,” ”covid” and ”trump” suggest dissapointment in the Govt. in power then.
5	False	Positive	#JoeBiden	Support Joe	Words like ”please,” ”Joe,” ”win” suggest healthy support to Biden but remaining words don’t add much meaning*
6	False	Positive	#DonaldTrump	Trump winning again	The words ”winning,” ”again,” ”peace”, ”best,” ”president” suggest prejudiced partisanship in Trump supporters
7	False	Negative	#JoeBiden	Schemers	The words ”corruption,” ”money,” ”scandal,” ”liars” imply allegations of Biden’s involvement in corruption.
8	False	Negative	#DonaldTrump	Govt doesn’t care	The words ”Liar,” ”don’t care,” ”filthy,” ”scum” suggest growing hatred towards Trump Govt
9	False	Negative	Overall	Is vaccine safe?	The words ”covid,” ”China,” ”vaccine” suggests prevalence of misinformation regarding Covid-19 vaccine.

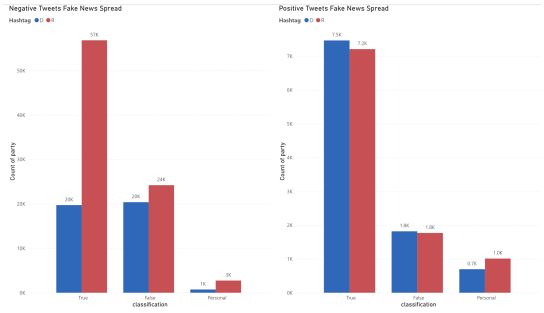

Figure 6: News Spread per Hashtag per Sentiment - Negative sentiment tweets (Left) are four times more than positive ones (Right), with equal distribution of true and fake news for positive sentiments. #DonaldTrump has a higher volume of true negative sentiment, suggesting significant public disappointment based on facts rather than falsehoods.

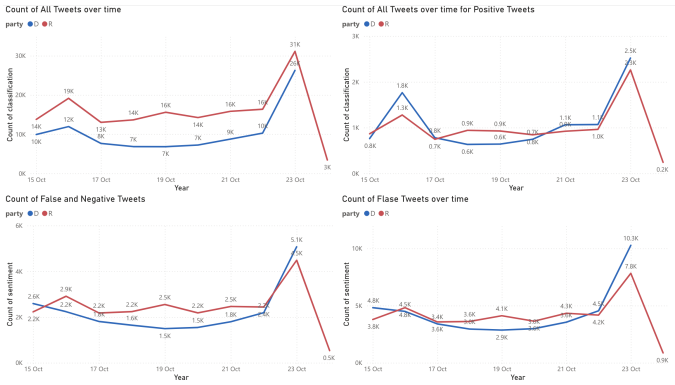

Figure 7: Graphs with combinations of Classification and Sentiment - Blue is #JoeBiden and Red is #DonaldTrump Hashtag tweet count. The first one (top Left) depict trends in tweet volume for each hashtag. Second one (Top Right) depict trends of all positive tweets per hashtag. Third (Bottom Left) depicts trends False and Negative Tweets per hashtag. Forth (Bottom Right) depicts trends in Fake news for each Hashtag. Both show similar Positive trend while the volume of false and negative tweets diverges significantly, with Republicans consistently posting higher volumes than Democrats.

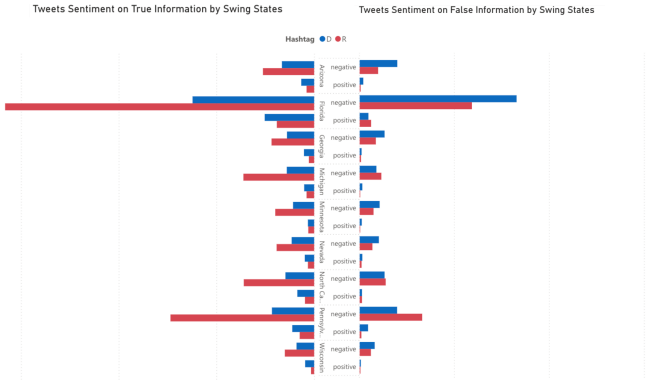

Figure 8: Sentiment by Swing State per Hashtag per Classification (Scale 0-1600) - The above graph gives insights on swing states and their divergence of Fake classification based on sentiment. Florida and North Carolina were the only swing states Republicans won, both of which had the highest negative spread in true tweets. Conversely, Democrats won in all other swing states, which featured either more Anti-Republican tweets or similar counts for both parties in both classifications.

It also identified distinct strategies in Twitter use by political groups, with anti-democratic narratives focusing on personal attacks, and pro-democratic ones promoting unity and voter mobilization. These results underscore the effectiveness of AI in analyzing online misinformation and sentiment, emphasizing the importance of ongoing research to enhance understanding of social media’s influence on public opinion and electoral outcomes.

Conclusion

The study on misinformation during the 2020 U.S. Presidential Election, utilizing advanced AI models like Few-shot Llama 3 and RoBERTa, revealed complex factors such as user behavior and regional demographics impacting Twitter’s political discourse, rather than straightforward partisan bias. Key findings indicate a disconnect between Twitter sentiment and actual electoral results, particularly in swing states, highlighting the need for improved data analysis methods across different social media and electoral periods.